When are experiments corrupt?

Article (Accepted Version)

When are experiments corrupt?

Nimi Hoffmann

2018

Between 2010 and 2011, the UK's Department for International Development (DFID) funded a large-scale experiment on Kenyan schools. The policy experiment tested the effects of cutting teachers' salaries and hiring them on short-term contracts, but the intervention failed after sustained opposition from teacher unions and parent associations. This article critically revisits the narrative of how this experiment was designed, implemented and interpreted, finding evidence that the experiment violated empirical logic. It examines whether the theory of neopatrimonialism can explain the ways in which vested interests may have undermined the empirical logic of the experiment. By doing so, the analysis tests the explanatory utility of neopatrimonialism, and casts light on the conditions under which policy experiments in Africa may be anti-empirical.

Keywords: Randomized controlled trials, impact evaluation, social policy, corruption, neopatrimonialism

The relationship between research and policy in the global South has often been a vexed one. This is in part due to contestations over the role of donor-funded researchers from the global North in shaping the evidence base used to craft southern policies, such that northern researchers have become increasingly concerned to "avoid charges of paternalism" (Deaton and Cartwright 2016, 3). The dangers of paternalism are especially acute in contexts in which, as Mkandawire argues, "the political class has failed to establish a productive and organic rapport with their own intelligentsia."(Mkandawire 2000, 205) The consequence of this is that donors have often played a central role in commissioning and funding policy research from both northern and southern scholars. It is not surprising then, that the role of donor-funded research in shaping southern policies has been subject to critical scrutiny. Rather than strengthening the capacity of southern governments, critics argue that "the persistent interference by outside actors ... undermines the development of young into strong democracies as it puts governments at risk of losing control over their own policy agendas." (Koch and Weingart 2016, 2)

Yet such asymmetry of power may also have implications for the empirical logic of policy research. This paper examines an increasingly popular form of donor-funded research:

the use of randomised controlled trials to evaluate the outcomes of social policy interventions. Such *policy experiments* are typically conducted on large samples, are often characterised by an absence of informed consent, and tend to involve significant financial and political stakes. While scholars are only slowly beginning to explore the ethical implications of such experimental setups (Barrett and Carter 2010; Reddy 2012; Ravallion 2012; Baele 2013), this paper focuses on exploring how these power asymmetries may undermine the empirical logic of policy experiments.

To do so, I examine the publicly available information on a large-scale experiment to casualise teachers in Kenyan schools between 2010 and 2011, funded by DFID, the UK's Department for International Development (Bold et al. 2018). The acute political contestation over this intervention illustrates the ways in which the high stakes of a policy experiment can have a bearing on its design, implementation and interpretation, such that the empirical logic of the experiment is undermined. The aim of this paper is to consider whether the theory of neopatrimonialism can help explain the ways in which vested interests shape policy experiments. It is part of a broader exploratory project, which seeks to develop the preliminary framework for a more in-depth empirical investigation into the role of large-scale experiments in African education.

In the first section, I briefly explain how policy experiments work. In the second section I outline the Kenyan education experiment, setting out the authors' framing of the research question, design and interpretation of results. In the third section, I draw attention to critical gaps and omissions in the experimentalists' narrative. I show that, when considered against the available evidence at the time, the framing of the findings appears to violate empirical logic. In the fourth section, I discuss evidence that the Kenyan government and DFID had vested interests in the experimental outcomes. In the final section I consider whether neopatrimonialism can be used to theorise the connection between these vested interests and the experiment's anti-empirical logic. I argue that it cannot, and that the appeal to neopatrimonialism does not say anything interesting about the nature of the experiment's anti-empiricism. While this suggests that neopatrimonialism has limited explanatory potential, examining its applicability to policy experiments helps clarify the conditions under which policy experiments in Africa may be anti-empirical.

1. What are policy experiments?

Experiments seem to solve the problem of causal identification. Non-experimental methods typically seek to identify the hidden causes behind observed phenomena. But since the correlation of two events is insufficient to claim causation, the identification of causes using this logic is theoretically complex and subject to contestation. In contrast, experiments reverse this logic, and seek to identify the effects of causes (Holland 1986). To do so, they invoke a counterfactual conception of causation. Consider a world in which events *c* and *e* occur – the cause and its effect. We can imagine a counterfactual world in which, everything else being equal, had *c* not occurred, *e* would not have occurred either. Since it is impossible for this world and the counterfactual world to exist simultaneously, experiments aim to simulate the counterfactual world through randomisation (Collins, Hall, and Paul 2004). Randomly allocating subjects to two groups – one which receives the cause *c*, and another which does not receive *c*, approximates a situation in which both groups differ in no meaningful way from each other – on average – save for the difference in treatment. Comparing the average difference between the two groups then yields an estimate of the average treatment effect. Unlike non-experimental methods then, experiments seem to provide intuitive and uncontroversial evidence of the effects of causes, at least at the mean. For this reason, experiments have come to be widely viewed as the ideal methodology for causal identification across a range of disciplines in the medical and social sciences.

However, development economists have increasingly extended this view to claim that experiments should be used to design and evaluate the outcomes of social policy interventions in the global South. In a 2004 *Lancet* editorial titled "The World Bank is finally embracing science", Esther Duflo, an influential proponent of this view, writes, "Creating a culture in which rigorous randomized evaluations are promoted, encouraged, and financed has the potential to revolutionize social policy during the 21st century, just as randomized trials revolutionized medicine during the 20th." (The Lancet 2004, 732) I call these *policy experiments*.

Over the last two decades, there has been a dramatic increase in policy experiments in the global South. The number of policy experiments used in World Bank evaluations increased from a baseline of zero in 2000 to just over two thirds of all evaluations in 2010 (Bédécarrats, Guérin, and Roubaud 2017). By 2016, one of the most influential proponents of policy experiments in development economics, the Abdul Latif Jameel Poverty Action Lab (J-PAL) had posted 811 policy experiments in 74 countries, where the largest proportion had been conducted in African countries (Jatteau 2016 quoted in Bédécarrats, Guérin, and Roubaud 2017). This growing interest is reflected in a systematic review of all policy

experiments published in 'leading economic journals' between 2009 and 2014, where 63% of all studies were conducted in the global South (Peters, Langbein, and Roberts 2016, see appendix).

Such policy experiments tend to be characterised by ambitious aims: they seek to draw generalisable conclusions about the effects of policy interventions. To do so, they need to be conducted on large representative samples in the 'real world.' In the systematic review of policy experiments noted above, 89% of all papers sought to generalise their findings, where the majority of these had sample sizes ranging from thousands of individuals to millions of households (Peters, Langbein, and Roberts 2016).

This is not surprising given that generalisability has been a strong normative commitment underlying many policy experiments. Proponents of policy experiments in development economics have advocated for experiments to create "global public goods" that "can offer reliable guidance to international organizations, governments, donors, and NGOs beyond national borders." (Duflo, Glennerster, and Kremer 2007) To this end, they have argued for the establishment of an international body to scale them up, providing a universal database on what works and what does not work in development (Banerjee and He 2008). Thus, for instance, J-PAL claim that 300 million people have been reached by eleven scaled up interventions, from school-based deworming to cash transfers (J-PAL 2018).

Claims to generalisability, or external validity, have been critiqued on a number of grounds, not least that the effects in the sample are unlikely to be similar to the effects in the population due to general equilibrium and political economy effects (Heckman 1992; Moffitt 1992; Deaton 2010), while the perception that experiments are non-parametric and theory-free is inconsistent with claims to generalisability (Reddy 2012; Muller 2014; Deaton and Cartwright 2016).

Yet the real-world logistics of conducting experiments also matter. When an experiment seeks generalisability, it typically requires a large sample, and this in turn requires an implementing agency with a big footprint – a government, development agency, international NGO or large firm. Thus, 74% of the studies in the systematic review discussed above were implemented by one of these institutional actors (Peters, Langbein, and Roberts 2016, see appendix). While this allows scale, the involvement of large institutional partners also poses a significant empirical risk: the financial and political stakes involved in policy interventions may introduce a conflict of interests, so that experimentalists become biased in the experimental design or interpretation of results.

These risks are heightened by the fact that policy experiments have rapidly become a sizeable global industry, driven by what Bédécarrats et al. (2017, 18) describe as an "entirely new scientific business model." As with other new industries that have rapidly emerged in the 21st century, there is little regulatory oversight and limited financial transparency. As Bédécarrats et al. (2017, 6) note, although "[d]ata are not available to quantify the financial flows concerned … it is sure that they run to hundreds of millions of dollars. For example, although J-PAL does not publish financial reports, IPA's annual revenue rose from US$252,000 in 2003 to over US$39 million in 2015." (IPA, or Innovations for Poverty Action, is an NGO that J-PAL established to scale-up experiments). While there is an established body of research in the medical sciences investigating the causal structure underlying the strong association between industry funding of clinical trials and pro-industry results (Mirowski and Van Horn 2005; Norris et al. 2011; Bes-Rastrollo et al. 2013), scholars have only recently begun to examine the role of a conflict of interests in policy experiments (Faulkner 2014; Bédécarrats, Guérin, and Roubaud 2017).

There may also be a tension between policy experiments' aspirations for generalisability and participants' informed consent. If participants know that they are in an experiment, then they may behave differently than they would under non-experimental conditions. As a consequence, the outcomes of an experimental intervention might not scale up to an entire population, since the experimental participants and the population may respond differently to the same policy intervention. Yet as Peters et al. (2018, 48) note, the standard solution to this problem in medical research – assigning a placebo treatment to a third group – is not possible in most policy experiments. There is no obvious 'sugar pill' equivalent in education, for instance. As a consequence, participants in policy experiments are seldom made aware that they are in an experiment. In the systematic review noted above, only 20% of the policy experiments explicitly secured participants' informed consent (Peters, Langbein, and Roberts 2016).[1]

Moreover, even when participants are informed of a policy experiment, the large-scale, institutional nature of the study means that it may be difficult for them to opt out. It may be costly to leave the service of the implementing agency (such as switching schools or banks), or participants may be locked into the service (such as relying on social welfare), or the service may be the cheapest or most convenient option (such as using the closest hospital or supermarket). This occurs particularly in the context of cluster randomisation, where entire social clusters, such as schools or clinics, are randomly assigned to an intervention so as to minimise knowledge or resource spillovers within the social cluster (Raudenbush 2018). In

such cases, participants may be informed of the experiment, but have little or no ability to consent to being experimented upon.

The "suspension of the fundamental principle of informed consent", as Barret and Carter (2010, 520) point out, "raises the subtle but important distinction between treating human beings as willful agents who have a right to participate or not as they so choose, versus treating them as subjects to be manipulated for research purposes." A failure to grapple with this distinction, they argue, often leads "subjects, implementers or both to actively circumvent the research design, thereby undercutting the statistical raison d'être of the initial randomization." (2010, 519) As a consequence, they argue, the social distance between experimentalists and participants can inadvertently undermine the statistical rigour of an experiment. However, little attention has been paid to the ways in which such social distance can lead experimentalists to ask questions that are untethered to social contexts or to misinterpret results.

The risk of social distance is more acute for policy experiments in the global South. This is because they are overwhelmingly conceptualised and conducted by scholars who do not live in the societies they study. As an indication of this, I briefly analysed the bibliometric data from a recent systematic review of policy experiments in the global South: 87% of lead authors are at institutions in the United States or Western Europe, 9% are at the World Bank, and just 4% are located in Asia, while no first authors are located in Africa or South America (Peters, Langbein, and Roberts 2018).[2] The complete absence of first authors located in Africa is particularly stark, since experiments conducted in Africa constituted 52% of all the papers reviewed.

To be clear: this discussion does not imply that policy experiments on African societies are necessarily defined by the absence of informed consent and social distance. Rather, it suggests that those engaged in experimenting on African societies do so in a context fraught with the risk of failing to be sufficiently attentive to the social worlds they investigate, of failing to apply a rigorous empirical logic.

2. What was the Kenyan education experiment?

In thinking through what empirical logic means in practice, I find myself drawn again and again to a policy experiment on Kenyan schools conducted between 2010 and 2011 (Bold et al. 2013, 2018).[3] Backed by the United Kingdom's Department for International

Development (DFID), it tested the effects of cutting teachers' salaries and hiring them on short-term contracts. It sought to replicate the results of experiments in India and western Kenya, which found that NGOs that hired contract teachers substantially improved the test scores of learners at the primary schools they ran (Banerjee et al. 2007; Duflo, Dupas, and Kremer 2012). These experiments were thought to work since contract teachers were driven to improve their learners' results out of fear of dismissal. Moreover, because they were not civil service teachers, NGOs were able to hire them at substantially lower salaries. The existing experimental evidence suggested that contract teachers improved results while substantially lowering costs.

However, this experiment was different, since the policy intervention would occur in public schools run by the Kenyan government. The experimenters were concerned that the move from an NGO to a government implementing agency might affect the treatment. For this reason, they simultaneously tested the contract programme in public schools and NGO schools run by World Vision, a Christian evangelical organisation from the United States, funded in part by DFID (World Vision 2018). This meant that the sample was divided into three: 64 government schools, 64 World Vision schools, and 64 control schools from both sectors. At 96 of the treatment schools, contract teachers would be hired at KSH 5000 or $67/month, while 32 schools would hire contract teachers at KSH 9000 or $121/month. The 'low' salary treatment was equivalent to 25% of what an entry-level public teacher would earn, while the 'high' salary treatment was equivalent to 50% of an entry-level salary. All in all, 196 teachers and about 15,000 students were involved in the experiment. During the course of the experiment, the government announced plans to scale the pilot to hire 18,000 teachers affecting about 1.3 million children.

The outcome of the experiment suggested that the intervention failed: while test scores at NGO schools indicated substantial and statistically significant improvement, there was no observable difference in test scores for government schools. In fact, the relative effect of moving from NGO to government implementation was negatively correlated with test scores, where this negative correlation was statistically significant.

How did the experimenters explain the failure of the intervention? Bold and her co-authors point out that the Kenyan National Union of Teachers opposed the policy reform through legal action, street protests and a two-day national strike. As a consequence of their successful legal action, the government was forced to promise to absorb the 18,000 contract teachers in public schools into the civil service at the end of their contract. Nevertheless, teachers in the experiment were not formally covered in the negotiations between the union

and the government. Indeed, while all 18,000 contract teachers were eventually employed as permanent civil servants on a full salary, only 45% of teachers in their experiment were able to gain permanent employment. Bold and her co-authors, however, point to evidence that teachers in the government treatment were substantially more likely to believe that the union represented their interests and that they would eventually gain a permanent position as a result of the union's political and legal actions. Thus, on the basis of this belief, the threat of losing their jobs for poor performance was no longer credible for government teachers in the experiment. In contrast, NGO teachers were unlikely to believe that the union would affect them and did not believe they were similarly protected; consequently, they continued to teach under fear of dismissal. Indeed, the experimenters find union identification is negatively correlated with test scores, where this correlation is strong and statistically significant, and maps onto the differences between government and NGO outcomes.

On their account, the experiment therefore demonstrates the kinds of political economy problems that arise from scaling up interventions from NGOs to the state. "Large-scale policy interventions", they write, "are likely to provoke political economy reactions from groups whose *rents are threatened by reform*, creating an endogenous policy response that counteracts the objectives of reform – the 'seesaw effect' (Bold et al. 2013, 4–5, own emphasis). The implication of this analysis is that government teachers are an elite whose democratic opposition to a national policy is a form of rent-seeking, which "undermines" the intervention (this is mentioned several times, see pp. 27, 28, 29) The problem, they observe, is that "governments are frequently the only institutional actors capable of taking education policies to scale." (2013, 2) Scholars have subsequently taken this experiment as evidence that governments in the global South should outsource the provision of public schooling to NGOs or for-profit firms, which are accountable to their foreign donors and shareholders rather than teachers, and are therefore less prone to capture by these 'rent-seeking elites' (Mbiti 2016; Kwauk and Robinson 2016; Romero, Sandefur, and Sandholtz 2018) This has the virtue of ensuring scale, while minimizing the potential for opposition from teachers.

3. How did the experiment employ an anti-empirical logic?

Describing opposition to an unpopular policy as a form of rent-seeking is not new. As Mkandawire and Soludo (1999, 26) note in their synthesis of thirty empirical studies on structural adjustment programmes in Africa, the "demonization of local elites suggests … the need for a deus ex machina to devise and implement policies unencumbered with the clamour

of domestic politics. From there, it is only a short walk to the conclusion that foreign institutions should take the driver's seat as agents of restraint." However, what is distinctive about the Kenyan education experiment in this intellectual context is the claim that teachers constitute an elite, and that their use of the mechanisms of democracy – including the right to assemble, the right to withhold labour, and the right to legal recourse – constitutes an attempt to protect illegitimate rents. While this is part of a broader long-running discourse on public-sector teachers, particularly in the United States (see, for instance, the Wall Street Journal Editorial 2017), it is a fairly recent development in research on African teachers.

The concept of rent-seeking belongs to the broader conceptual family of corruption. It describes a situation in which an actor attempts to increase her share of wealth by manipulating the social or political environment, rather than by creating new wealth (Tullock 1967; Krueger 1974). It was initially used to describe situations in which a government creates an artificial monopoly through legislation, where aspiring monopolists then lobby the government to capture this monopoly. As Tollison (1982, 557) remarks, "It is this activity of wasting resources in competing for artificially contrived transfers that is called rent seeking." While Bold and her co-authors do not explain how this concept applies to teachers in their experimental study, in their subsequent analysis of the experimental dataset, they provide a loose definition of teachers' rent as "the gap between civil service wages and labor market outcomes for an ostensibly identical applicant." (Barton, Bold, and Sandefur 2017, 13) On this definition, any attempt by public sector teachers to increase their wages above those of private sector teachers for the same quality of teaching efforts constitutes a form of rent-seeking.

But the appeal to rent-seeking to explain why the experiment failed is a curious strategy for several reasons. First, unionised teachers did not explicitly protest to protect their own salaries, but instead stated that their protest was on behalf of non-unionised contract teachers who were new to the profession. In this regard, commentators in the Kenyan media described the teachers' strike as an act of sacrifice in solidarity with vulnerable contract teachers (see media reports from Education International 2012). While it is possible that some unionised teachers perceived the introduction of low-wage contract teachers as a long-term threat to their own incomes and job security, insofar as it opened the way for the low-wage casualisation of all teachers, it is also possible that some teachers had different motivations in place of, or in addition to, self-interest. That is, their motivations are not a given, but are instead a subject of empirical enquiry.

To understand why this is important, consider that the contracting of teachers represents a well-recognised threat to the institutional fibre of the education system, for, as the World Bank notes, contract teachers "can overcome the rigidity of existing teacher policies by establishing a parallel teacher corps or career stream with *different rules of the game* alongside the existing teacher stream." (Bruns, Filmer, and Patrinos 2011, 146, emphasis added) While proponents of contract teaching often see such institutional bifurcation as a positive outcome, stemming as it does from the introduction of private sector practices, a systematic review on the casualisation of teachers in the global South suggests a more complex picture (Chudgar, Chandra, and Razzaque 2014). It finds strong evidence that low-paid contract teachers are more likely to report job dissatisfaction and low morale relative to permanent teachers, and are more likely to leave teaching once better opportunities present themselves, leading to higher levels of teacher attrition. This destabilises the teacher market and leads to high levels of institutional churn. The authors caution that this institutional fracturing, coupled with teachers' "lack of future prospects, insufficient legal protection, complete reliance on local hiring, low pay, and limited support for teaching in challenging environments may send a message that teaching as a profession is not important and not valued." (Chudgar, Chandra, and Razzaque 2014, 156) Against this backdrop, it is plausible that at least some union teachers might have been concerned that the introduction of low-wage contract teachers could threaten the long-term sustainability and flourishing of the education system. Indeed, it is surprising that Bold and her co-authors did not consider this explanation, given that their experimental study found offering "a higher salary had significant positive effects on schools' ability to recruit and retain contract teachers." (Bold et al. 2013, 20)

Related to this, union teachers may have been concerned with the financial wellbeing of contract teachers. Hiring contract teachers at the low salary of KSH 5000 or $67/month in 2011 meant that they would have a daily income of $4.70 per day in PPP (purchasing price parity) dollars.[4] To understand the significance of this, assume that two married teachers were each employed at this contract wage, and had three children; each person's income in the family would then be $1.88 per day in PPP dollars. This falls below the World Bank's poverty line of $1.90 per day.[5] Put in perspective, the cost of a minimally healthy food basket for one person in Kenya at this time ranged from $1.27 to $2.17 per day on a PPP basis (World Food Programme 2012). In most cases then, teachers would be unable to feed themselves sufficiently on their contract wages, and would not have income for any other living costs, such as housing, electricity, water, transport or healthcare. Even if a teacher did

not have any child or adult dependents, a daily wage of $4.70 would be unlikely to cover these additional living costs. While teacher surveys across a number of African countries indicate that permanent teachers typically do not earn enough to cover their monthly household expenses (Koomson, Afful, and Villano 2017), such that over one-third of respondents in teacher surveys indicate that teachers go to school hungry (Bennell and Akyeampong 2007) the need to smooth consumption is likely to be far more acute for contract teachers, who receive substantially lower wages.

These low wages may introduce perverse incentives into education systems, since teachers are forced to supplement their income in order to survive. Teacher can do so by taking credit, getting a second job (moonlighting), or engaging in illicit financial activities to meet their basic living needs. The financial distress of indebtedness is strongly linked to high levels of teacher attrition (Koomson, Afful, and Villano 2017), which in turn destabilises school functioning in the long-term. Teachers who moonlight are more likely to be absent from school or not engaged in teaching while at school, where this is strongly correlated with lower learning outcomes (Chaudhury et al. 2006). And poorly paid teachers may illicitly require parents to pay for examinations or additional tutoring in order for their children to pass or attend school (Osler 1997; Hofer 2003; Sayed et al. 2018). Where teachers do not levy illicit fees on parents they may instead attempt to extract additional money from the education system through corrupt hiring practices. Across a number of different contexts, the introduction of contract teachers has been marked by illicit hiring practices, such as bribes for fast-tracking contract teachers into regular teaching (Chudgar, Chandra, and Razzaque 2014). In Kenya, an earlier contract teacher experiment found that, without oversight, 31% of additional teachers hired on contract basis were relatives of civil-service teachers already working at that school (Duflo, Dupas, and Kremer 2015).

Consequently, the introduction of low-wage contract teaching does not only have negative implications for the basic wellbeing of contract teachers, it also has implications for the wellbeing of learners, parents and the education system as a whole. It is not surprising then that parents' associations and non-governmental organisation joined teacher unions in protesting against the introduction of low-wage contract teachers (Education International 2012). Yet Bold and her co-authors fail to mention this information in their study. This is significant, for parents' rejection of the contract teacher programme casts doubt on the claim that opposition to the intervention was a form of rent-seeking, insofar as parents were plausibly motivated out of concern for the potential negative effects on their children or the education system as a whole, rather than an attempt to protect teachers' salaries.

There is a further omission of critical information from the explanation. Bold and her co-authors note that teachers sought legal action in order to compel the Kenyan government to hire contract teachers as permanent staff with the same wages and benefits as other teachers. However, the authors fail to mention that the court ruled that employing a subset of teachers on contract and at much lower wages violated constitutional provisions for equal and fair employment and contravened the Teacher Service Act, which makes employment of government teachers on contracts illegal (Education International 2012). In other words, the teacher union argued that equal work deserves equal pay and conditions of employment. This argument is structured around an anti-discrimination logic, as the Kenyan courts recognised. It is surprising that the authors fail to mention this information, as it was central to the teacher union's successful legal action.

It is not obvious that introducing different pay for the same work is a way of improving efficiency, rather than a form of discrimination. It is difficult to see how this claim could be sustained, for instance, in the context of lower pay for black people or women in the labour market.[6] The apparently counter-intuitive implications of such a claim therefore shift the burden of proof to the experimentalists, who must show that teachers' attempts to prevent unequal pay for equal work were a form of rent-seeking and not a form of anti-discrimination procedures. Yet they provide no evidence for this claim. Instead, the experimentalists appear to use rent-seeking as an *a priori* proposition. *A priori* propositions are those that need no empirical evidence; they are deductively true. An example of this is the statement that bachelors are unmarried men. But the claim that Kenyan teachers' opposition to a policy of different pay for the same work is a form of rent-seeking is not deductively entailed by their status as Kenyan teachers or their opposition to a government policy. It may turn out to be the case that unequal pay for the same work is an efficient market outcome, rather than a form of labour market discrimination, but this requires empirical and theoretical justification; it cannot be treated as a self-evident truth.

This discussion suggests that the appeal to rent-seeking to explain experimental failure is an example of anti-empirical logic on several different fronts. First, it assumes teachers' motivations, rather than subjecting them to empirical enquiry. Second, it disregards the well-documented negative effects from contract teacher programmes, which would plausibly motivate public-spirited opposition to the programme. Third, it omits critical empirical evidence from the intervention that appears to conflict with the appeal to rent-seeking. And fourth, it assumes that teachers' opposition to a programme of different pay for

the same work was a form of rent-seeking, effectively taking this to be a deductively true statement.

The anti-empirical nature of the experimental study becomes clearer when we consider the social costs of the experiment. It not only failed to produce an observable improvement in children's test scores, but catalysed major volatility in the education system due to the illegal and unconstitutional nature of the programme. In doing so, it plausibly harmed the already fragile relationship of trust between teachers and government that is vital to ensuring a stable and well-functioning education system. Empirically-minded scholarship would acknowledge that the experiment failed because the treatment was very likely damaging, and would therefore seek alternative treatments. This, however, was not the approach in the Kenyan experiment. Bold and her co-authors elided the court's findings that the programme was illegal and unconstitutional, and did not acknowledge that the treatment was likely damaging to the education system. Instead, teachers' *objections* to the damaging treatment were taken to be the reason for experimental failure. This suggests that the experimentalists were committed to the view that the intervention was an inherently positive one, which was distorted by teachers' agency. The solution, as one of the authors of the Kenyan experiment has subsequently argued, is to prevent teachers from exercising their agency in future experiments by placing control of the public education system in the hands of foreign actors (Romero, Sandefur, and Sandholtz 2018).

4. Evidence of vested interests in the experimental outcomes

The preceding section suggests that the Kenyan education experiment was marked by an anti-empirical logic, insofar as it was marked by *a priori* reasoning and elided or disregarded well-known evidence at odds with the conclusions of the experimentalists. This section considers one potential factor underlying this anti-empirical logic: evidence that the Kenyan government and DFID had vested interests in the experimental results.

First, the Kenyan government likely viewed the contract teacher programme as a political necessity. To understand why, consider that the intervention was undertaken in a context marked by systematic under-funding of public education. The government had first introduced the principal of free primary education in the decade following independence, but had been forced to cut education spending substantially and introduce tuition fees under structural adjustment in 1989. In 2001, the government, under Daniel arap Moi's

administration, announced that free primary education would be reintroduced within two years. When the administration backtracked on this commitment in 2002, mass protests by teacher unions, parents, and NGOs forced the government to re-commit to free primary education (Mukudi 2004). That year, the National Alliance Rainbow Coalition managed to unseat Moi after 24 years of rule, partly on the back of promising free primary education for all. Within a year of abolishing primary school fees in 2003, the enrolment rate increased by 22.3% (Oketch and Rolleston 2007). While the Kenyan government raised its education budget in 2003-04 by 17.4% and was strongly supported by donor funding, it was estimated that government would require an increase of at least 91% to the budget, or 0.8% of the GDP, to finance a net enrolment rate of 94% (Vos et al. 2004). However, funding constraints meant that this budgetary shortfall was never met. By 2006, the shortfall in the government's education budget stood at 55%, or $745 million, and while the government borrowed $80 million from the IDA and $388 from donors including DFID, it still had a shortfall of $277 million (Omwami and Keller 2010).

Given the politically unfeasibility of scrapping free primary education, and under pressure from international agencies and donors, the government responded to the budgetary shortfall by implementing a hiring freeze on teachers (Amutabi 2003). By 2011, the Ministry of Education reported a shortage of 61,000 public sector teachers relative to its target of a 40:1 learner-teacher ratio (Bold et al. 2013). However, the hiring freeze on teachers appears to have been accompanied by a marked drop in education quality, as teachers battled with overcrowded classrooms and a lack of teaching and learning resources. This is likely reflected in the difference in growth between the net and gross enrolment rate in primary schools. The net enrolment rate increased from 74% in 2004 to just under 80% in 2007, while the gross enrolment rate climbed more steeply from 100% to 112% over the same period, where this difference plausibly indicates high levels of learners repeating school (Omwami and Keller 2010). The apparent decrease in education quality has in turn driven demand for low-fee private schools. In this respect, an analysis of a panel survey of 718 rural Kenyan households conducted in 2004 and 2007 finds that the probability of attending private schools increased as the average pupil–teacher ratio of public schools increased in the community (Nishimura and Yamano 2013). The authors note that, between 2002 and 2005, the number of private schools increased by more than 400%, while the number of public schools increased by just 1.6%.

Within this context, there have clearly been strong political imperatives to find cost-effective means for increasing the number of teachers in public schools. The political

sensitivity of the education budget is in part reflected by the fact that the Minister of Education at the time, Uhuru Kenyatta, was simultaneously the Minister of Finance. It may also explain why the "scale-up of the national program occurred before the pilot was completed due to political pressure from outside the Ministry of Education." (Bold et al. 2013, 8) And despite the clear evidence that the contract teacher programme was an empirical and political failure, in 2018 the government announced plans to hire 68,000 teachers as "interns" at half the wage of civil service teachers (Wanzala 2018).

Another set of vested interests may be located in the funding agency, DFID. In 2010, DFID announced that it would spend £36 million to promote private education and health services during 2011-2015. Although this was a fraction of DFID's aid spending, observers have noted that its reach has been much greater: "it is 'helping' developing countries revise policies to facilitate the private provision of services and making such approaches, which are politically controversial, more acceptable." (Curtis 2015, 6) This in turn paved the way for the launch of its 2011-15 operational plan, which signalled a shift from funding public education services to investing in private education services as part of a new strategy to promote education exports and increase its revenue (DFID 2011). At the same time, it also announced a £355 million education project, *The Girl's Education Challenge*, which aims to promote non-state (for profit and NGO) provision of education. The Department for Business, Innovation and Skills, together with UK Trade and Investment, concurrently set up a new body, UKTI Education to identify "high value commercial education opportunities" for UK businesses (UKTI Education 2011). Its "priority markets" include East Africa and India, areas where DFID is very active in education. The following year, DFID launched its Impact Fund, a for-profit investment fund worth £75 million set up with the aim of "foster[ing] a revolution in private investment into projects to improve the health, education and future chances of more than five million people across Africa and Asia … British aid will provide financial capital and specialist advice to transform health and education services." (UK Government 2012)

While the market for private educations services is estimated by market analysts to be $1.5-2.0 billion in the next five years, the problem as the analysts observe, is one of scale – companies servicing poor populations require scale in order to reach profitability (Caerus Capital 2017). In this respect, one of DFID's flagship investments is Bridge International Academies, an American firm that operates a chain of for-profit schools in a number of countries, including Kenya. Bridge's business model relies on hiring unqualified community members on low contract wages to teach scripted curricula via tablets. (Kwauk and Robinson

2016) It currently has 100,000 students, but requires 500,000 students to break even. It eventually aims to educate 10 million students across twelve countries by participating in public private partnerships with African governments that would allow it to take over the management of the public education sector (Anderson 2018).

This suggests that DFID has vested financial interests in promoting large-scale private education services in Kenya, particularly through the mechanism of public private partnerships. However, as discussed above, it appears that mass demand for private schools in Kenya is at least in part a function of deteriorating public schools. This means that a strong state-managed public education system is likely incompatible with a large-scale low-cost private education system or a privately-managed public system. Thus, any interest DFID might have in strengthening a state-managed public education system is complicated by its financial interests in expanding the private education sector.

This is significant, because DFID not only funded the Kenyan education experiment, but its chief economist at the time, Stefan Dercon, also helped to "conceive" it (Bold et al. 2013, 1 footnote) Furthermore, DFID then interpreted the findings of the study as grounds for further private sector intervention in public schools (Ashley et al. 2014). This appears to be a form of circular reasoning: the experiment was designed to take a feature of the private sector – low-wage, unprotected and casualised labour – and apply it to the state sector on the grounds that the private sector is superior. The failure of the intervention, and the major volatility it catalysed in the public sector, were then taken as grounds for further private sector intervention. Yet, even if the private sector treatment had succeeded in the experiment, this would plausibly have been taken as grounds for increasing private sector intervention in public schools. It appears that the experimental design was such that, whatever its outcomes, it would support greater private sector involvement in schooling.

Taken together, the discussion suggests that the experimental design and interpretation of results were favourable to the interests of the Kenyan government and DFID. Yet the causal processes by which these actors' vested interests may have influenced the experiment remain unclear. This is in part because there is insufficient empirical detail. Outside of an acknowledgement that DFID's chief economist at the time helped to design the experiment, the genesis of the experiment, the reasons for scaling up the experiment before it was completed, and the details of the relationship between the Kenyan government, DFID and the experimentalists remain hazy. However, as Cartwright (1999) has persuasively argued, it is also difficult to discern causal processes without some kind of theoretical buttressing. As a way of laying the foundations for empirical work on the topic, how might

one theorise the causal relationship between vested interests and anti-empiricism in experiments?

5. Is this a neopatrimonial project?

The entanglement between profit, politics and experimentation raises the prospect of collusion for political or financial profit at the expense of vulnerable children and teachers. The apparent divergence in interests between the elites who funded, designed and implemented the experiment, and the interests of the teachers and children who were experimented upon suggests a particular kind of collusion: neopatrimonialism. This concluding section considers the extent to which the theory of neopatrimonialism might help explain the connection between vested political and financial interests on the one hand, and experimental outcomes on the other.

Neopatrimonialism is a term that is often invoked to explain perceived failures in African societies. Its contemporary usage is steeped in the ethnographic language of tribe, ethnicity, big men, and patronage (Mkandawire 2015). But when stripped of this ethnographic language, it seems to describe a society characterised by profound power inequalities between an elite and a citizenry who are hermetically sealed off from each other. The interests of these two groups invariably conflict, and while citizens may try to get these interests to coincide by becoming clients of elites, they do so in vain. Elites always act in their own interests and dispossess citizens. In their single-minded pursuit of their own interests, however, they must discount any evidence for or against the broader social benefits of a given policy. That is, they must employ an anti-empirical logic.

The problem with this account, as a number of scholars have pointed out, is that it is difficult to observe neopatrimonialism in action, not least because it does not accurately describe any given society (Mustapha 2012; Mkandawire 2015). One might respond that neopatrimonialism is a model that necessarily abstracts away empirical complexity. Still, this response does not address the empirical question of how one identifies neopatrimonialism at work. This is because neopatrimonial relations involve individual interests, which are latent or unobservable. A number of scholars working in the neopatrimonialism framework therefore invoke the observed phenomenon of anti-empirical logic to deduce that neopatrimonialism must be at work – a kind of revealed preference theory of corruption. Perhaps the most well-known example of this is Van de Walle's (2001) argument that African elites ignored the evidence that structural adjustment programmes would result in

large social benefits and therefore did not implement adjustment programmes; this was taken as evidence that structural adjustment would endanger their interests, and therefore, that neopatrimonialism was at work.

Using this reasoning then, the evidence of anti-empirical logic by an elite counts as an example of neopatrimonialism. In this regard, the Kenyan experiment was funded by a donor elite, implemented by a political elite, and conducted by an academic elite. The academic elite invoked anti-empirical logic to explain their subjects' resistance to the policy intervention. The political elite disregarded the empirical evidence in scaling up the project. And the donor elite have continued to fund similar projects despite experimental failure in Kenya (Muralidharan and Sundararaman 2013).The experiment therefore meets the conditions for neopatrimonialism. Notice that this assessment does not require further information about the experimenters' actual interests or the ways in which they may be entangled in institutions with vested interests in observing particular results. Instead, neopatrimonialism simply requires evidence of anti-empiricism among elites.[7]

Yet, considered carefully, categorising the Kenyan experiment as a form of neopatrimonialism only obscures the analysis, rather than clarifying it. To see why, consider that there are a number of ways of analysing the anti-empirical logic of scholars working on African societies. Perhaps the most common analyses describe the reasoning underlying anti-empiricism. Thus, for instance, Mamdani (1990) argued that American scholars working on Africa during the Cold War era were stuck in an intellectual cul de sac, since they had borrowed theoretical paradigms developed elsewhere and then tried to force reality to fit these borrowed theories rather than developing theories to fit reality. One can call this the *argument from mimicry*. These low scholarly standards, critics argued, were in part a consequence of academic work being constrained by the foreign policy objectives of the Cold War, such that political patronage impeded the development of rigorous scholarship and intellectual dissent in American scholarship on Africa (Guyer 1990; Berger 1997; Zeleza 2003). There is some evidence that the Kenyan experiment may fit this picture, in the form of experimenters' attempts to transpose the programme logic from projects that were run by international NGOs and for-profit schools to projects that were run by governments. The key difference between the two is that the former are shaped by an anti-democratic institutional logic, insofar as foreign donors and investors are their primary constituency, while the latter are bound by a more democratic institutional logic, insofar as citizens are their primary constituency. In this context, it is possible that DFID's financial interests in private schools

may have prompted the experimenters' interest in transposing an intervention from the private sector into the public sector.

Related to this is the *argument from ignorance*: scholars who attempt to engineer social reforms without understanding the society at hand. This is perhaps one of the most common critiques of structural adjustment programmes on the continent (Mkandawire and Soludo 1999). Yet, as Olukoshi (2006, 540) argues, the epistemic constraints on northern scholars in this context derived not only from their relationship to power, but also from their intellectual dependence on African scholarship, such that the decline in African intellectual communities during the 1980s signalled a concomitant decline in African studies, particularly in the United States. In this regard there is some evidence of epistemic constraints on the Kenyan experiment. A review of the authors' résumés indicates that only the lead experimentalist, Tessa Bold, had conducted education research prior to the experiment and none of the Kenyan experimentalists were education economists or had training in education research as a discipline. This perhaps explains why the experimental design does not consider that the chronic underfunding of education, which began under structural adjustment, has undercut the material basis of the Kenyan education system. This has led to the persistent non-payment of teachers, a collapse in wages that has yet to recover, a freeze on hiring, and low rates of investment in teacher education, libraries and school feeding programmes – all elements of a quality education system (Stromquist 1999; Klees, Samoff, and Stromquist 2012; Stromquist, Klees, and Lin 2017). Given this context, it is difficult to understand why the experimental design sought to introduce further precarity and impoverishment into the education system, rather than testing the effects of strengthening its financial and administrative base, something which a number of education economists have called for (Mukudi 2004; Omwami and Keller 2010; Ron Balsera, Klees, and Archer 2018).

But policy experiments in Africa are also conducted against a historical background of colonial experimentation on Africans. In the 1930s, colonial scientists conducted wide-ranging experiments in British colonies that provided the underpinning for contemporary tropical medicine, agriculture and ecology. These experiments were extensive and influential enough for historians to call Africa a "living laboratory" (Celik 1997; Tilley 2011).[8] They were often harmful and dehumanising. By refusing to allow participants to opt out of the experiment, and by treating their dissent as a form of rent-seeking corruption, the experiment objectified Kenyan participants in ways that echo colonial science. Here, anti-empirical logic manifests in experimenters' lack of attentiveness to the full humanity of participants, which plausibly stemmed, at least in part, from the deep power asymmetries between the

experimenters and their subjects. I call this the *argument towards colonial experimentation:* when experimentalists objectify subjects such that experimentalists do not attend to the meanings and motivations of subjects' actions.

These arguments do not imply that donor-funded policy experiments are simply a continuation of colonial and cold war intellectual histories in Africa. Rather, these arguments draw our attention to the ways in which the anti-empirical reasoning in policy experiments may be shaped by specific political, economic and historical conditions. Common to all three arguments is a profound asymmetry of power, both political and financial, which helps to protect elite experimenters from empirical reality. In the resulting echo chamber, there may have been little to stop those in the Kenyan experiment from interpreting their own failure as an outcome of rent-seeking by unruly African teachers. If this is correct, then it is unlikely that anti-empirical logic is a result of experimentalists simple refusing to be attentive to the social world they investigate. Instead, it is more likely that the conditions under which they conduct experiments place significant constraints on their capacity to be sufficiently attentive. For as Tamale and Oloka-Onyango (1995, 694) argue, scholars who are too close to power may become "inoculated from the very processes and peoples being examined."

Given this analysis, what work does the concept of neopatrimonialism do when trying to understand anti-empirical logic? The concept cannot be used to distinguish between the different forms of anti-empiricism, nor can it be used to explain them. Moreover, invoking neopatrimonialism to describe the Kenyan experiment flattens social theorising by rendering experimenters' intentionality and behaviour irrelevant. As Mkandawire (2015) has argued, this is because the concept is poorly defined, empirically fuzzy and analytically unhelpful. In this regard, the concept functions best as a rhetorical tool: it is useful in signalling moral approbation, but unhelpful in empirical research. The implication of this is that, no matter how tempted one is, one should avoid invoking neopatrimonialism and employ well-defined concepts instead. The problem, as Mkandawire (2015, 601) notes, is that when it comes to claims about neopatrimonialism in African societies, "easy acceptance by researchers and policymakers of the school's claims in the absence of empirical evidence suggests that there are strong preconceptions and prejudices about African politics that are unlikely to be dispelled by a more accurate measurement of the phenomenon in question." Turning the logic of neopatrimonialism on northern scholarship is one strategy for demonstrating the empirical paucity of the concept. As this analysis has sought to demonstrate, using a northern-funded and designed experiment as a case study provides a clearly-bounded and concrete illustration of the way in which neopatrimonialism obscures rather than clarifies the object of analysis.

In closing, the analysis suggests that policy experiments are neopatrimonial when they deploy anti-empirical logic, but it turns out that this does not mean much beyond signalling moral displeasure. If neopatrimonialism is an impractical tool for investigating a very concrete, bounded social phenomenon, like an experiment, then it is likely to prove equally useless when thinking about highly complex societies with porous boundaries. While neopatrimonialism may not be a useful theoretical framework, examining its applicability to policy experiments casts light on the ways in which the ambition of policy experiments can engender power asymmetries that constrain experimentalists' capacity to attend to their empirical context. This suggests a more nuanced picture. Rather than experimentalists simply ignoring or explaining away inconvenient results, it seems that the very ambition to greater empirical validity is a constraint on empirical validity. If this is true, then the ways in which experimentalists understand and navigate their constraints is critical to their capacity to employ an empirical logic. It follows from this that experimentalists' agency is of central theoretical importance in investigating the relationship between vested interests and policy experiments.

References

Amutabi, M.N. 2003. 'Political Interference in the Running of Education in Post-Independence Kenya: A Critical Retrospection'. *International Journal of Educational Development* 23 (2): 127–44. https://doi.org/10.1016/S0738-0593(01)00055-4.

Anderson, Jenny. 2018. 'The Controversial Silicon Valley-Funded Quest to Educate the World's Poorest Kids'. *Quartz Africa*, 22 January 2018. https://qz.com/1179738/bridge-school/.

Ashley, Laura Day, Claire Mcloughlin, Monazza Aslam, Jakob Engel, Joseph Wales, Shenila Rawal, Richard Batley, Geeta Kingdon, Susan Nicolai, and Pauline Rose. 2014. 'The Role and Impact of Private Schools in Developing Countries–Bibliography and Literature Reviews'. London: DFID.

Baele, Stéphane J. 2013. 'The Ethics of New Development Economics: Is the Experimental Approach to Development Economics Morally Wrong?', 42.

Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden. 2007. 'Remedying Education: Evidence From Two Randomized Experiments in India'. *Quarterly Journal of Economics* 122 (3).

Banerjee, Abhijit, and Ruimin He. 2008. 'Making Aid Work'. In *Reinventing Foreign Aid*, edited by William Easterly, 47–92. Cambridge, MA: MIT Press.

Barrett, Christopher B., and Michael R. Carter. 2010. 'The Power and Pitfalls of Experiments in Development Economics: Some Non-Random Reflections'. *Applied Economic Perspectives and Policy* 32 (4): 515–48. https://doi.org/10.1093/aepp/ppq023.

Barton, Nicholas, Tessa Bold, and Justin Sandefur. 2017. 'Measuring Rents from Public Employment: Regression Discontinuity Evidence from Kenya'. RISE-W P-17/ 01 5. Research on Improving Systems of Education (RISE).

https://www.riseprogramme.org/sites/www.riseprogramme.org/files/publications/RISE_WP-015_Barton%2C%20Bold%20%26%20Sandefur.pdf.
Bédécarrats, Florent, Isabelle Guérin, and François Roubaud. 2017. 'All That Glitters Is Not Gold. The Political Economy of Randomized Evaluations in Development'. *Development and Change* 0 (0): 1–28. https://doi.org/10.1111/dech.12378.
Bennell, Paul, and Kwame Akyeampong. 2007. 'Teacher Motivation in Sub-Saharan Africa and South Asia'. Educational Papers. London: DFID.
Berger, Iris. 1997. 'Contested Boundaries: African Studies Approaching the Millennium Presidential Address to the 1996 African Studies Association Annual Meeting'. *African Studies Review* 40 (2): 1. https://doi.org/10.2307/525154.
Bes-Rastrollo, Maira, Matthias B. Schulze, Miguel Ruiz-Canela, and Miguel A. Martinez-Gonzalez. 2013. 'Financial Conflicts of Interest and Reporting Bias Regarding the Association between Sugar-Sweetened Beverages and Weight Gain: A Systematic Review of Systematic Reviews'. *PLoS Medicine* 10 (12): e1001578.
Bhorat, Haroon, Karmen Naidoo, Morne Oosthuizen, and Kavisha Pillay. 2015. 'Demographic, Employment, and Wage Trends in South Africa'. WIDER Working Paper 2015/141. Helsinki: United Nations University World Institute for Development Economics Research.
Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, Alice Ng'ang'a, and Justin Sandefur. 2013. 'Scaling up What Works: Experimental Evidence on External Validity in Kenyan Education'. Working Paper 321. Centre for Global Development.
Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, Alice Ng'ang'a, and Justin Sandefur. 2018. 'Experimental Evidence on Scaling Up Education Reforms in Kenya'. *Journal of Public Economics*, 59.
Bruns, Barbara, Deon Filmer, and Harry Anthony Patrinos. 2011. *Making Schools Work: New Evidence on Accountability Reforms*. The World Bank.
Caerus Capital. 2017. 'The Business of Education in Africa'. Washington, D.C.: Caerus Capital. https://www.avca-africa.org/media/1739/the-business-of-education-in-africa-2017.pdf.
Cartwright, Nancy. 1999. *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.
Celik, Zeynep. 1997. *Urban Forms and Colonial Confrontations: Algiers under French Rule*. Berkeley: University of California Press.
Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Halsey Rogers. 2006. 'Missing in Action: Teacher and Health Worker Absence in Developing Countries'. *Journal of Economic Perspectives* 20 (1): 91–116. https://doi.org/10.1257/089533006776526058.
Chudgar, Amita, Madhur Chandra, and Ayesha Razzaque. 2014. 'Alternative Forms of Teacher Hiring in Developing Countries and Its Implications: A Review of Literature'. *Teaching and Teacher Education* 37 (January): 150–61. https://doi.org/10.1016/j.tate.2013.10.009.
Collins, John David, Edward Jonathan Hall, and Laurie Ann Paul. 2004. *Causation and Counterfactuals*. Massachuesets: MIT Press.
Curtis, Mark. 2015. 'Profiting from Poverty - Again'. London: Global Justice Now.
Deaton, Angus. 2010. 'Instruments, Randomization, and Learning about Development'. *Journal of Economic Literature*, 424–455.
Deaton, Angus, and Nancy Cartwright. 2016. 'Understanding and Misunderstanding Randomized Controlled Trials'. NBER Working Paper 22595. National Bureau of Economic Research.
DFID. 2011. 'Operational Plan 2011-15: Private Sector Department'. London: DFID.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2012. 'School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools'. MIT Department of Economics Working Paper Working Paper 12-07. Cambridge, MA: MIT.

———. 2015. 'School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools'. *Journal of Public Economics* 123: 92–110.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2007. 'Using Randomization in Development Economics Research : A Toolkit'. In *Handbook of Development Economics*, edited by T. Paul Schultz and John A. Strauss, 4:3895–3962. Amsterdam: Elsevier.

Education International. 2012. 'How KNUT Got Rid of Contract Teachers'. Education International. 2012. https%3A%2F%2Fei-ie.org%2Fen%2Fdetail%2F1920%2Fhow-knut-got-rid-of-contract-teachers.

Faulkner, William N. 2014. 'A Critical Analysis of a Randomized Controlled Trial Evaluation in Mexico: Norm, Mistake or Exemplar?' *Evaluation* 20 (2): 230–43. https://doi.org/10.1177/1356389014528602.

Guyer, Jane. 1990. 'A Comment on Mamdani'. *CODESRIA Bulletin*, no. 3: 10.

Hailey, Lord. 1938. *An African Survey: A Study of Problems Arising in Africa South of the Sahara*. London: Oxford University Press.

Heckman, James J. 1992. 'Randomization and Social Policy Evaluation'. In *Evaluating Welfare and Training Programs*, edited by Charles Manski and Charles Garfinkel, 547–70. Cambridge, Massachusetts: Harvard University Press.

Hofer, Katharina. 2003. 'The Role of Evangelical NGOs in International Development: A Comparative Case Study of Kenya and Uganda'. *Africa Spectrum* 38 (3): 375–98.

Holland, Paul W. 1986. 'Statistics and Causal Inference'. *Journal of the American Statistical Association* 81 (396): 945–60.

J-PAL. 2018. 'Scale-Ups'. The Abdul Latif Jameel Poverty Action Lab. 2018. https://www.povertyactionlab.org/scale-ups.

Klees, Steven J., Joel Samoff, and Nelly P. Stromquist, eds. 2012. *The World Bank and Education: Critiques and Alternatives*. Vol. 14. Comparative and International Education: A Diversity of Voices. Rotterdam, Boston, Taipei: Sense Publishers. https://books.google.co.za/books?hl=en&lr=&id=DYxgM1im9o0C&oi=fnd&pg=PR5&dq=%22structural+adjustment%22+%22nelly+stromquist%22&ots=k1uiOMHw_u&sig=U1ZB8YYUWs_724eTq1xGcTEX5Yk#v=onepage&q=stromquist&f=false.

Koch, Susanne, and Peter Weingart. 2016. *Delusion of Knowledge Transfer: The Impact of Foreign Aid Experts on Policy-Making in South Africa and Tanzania.* Stellenbosch: African Minds.

Koomson, Isaac, Benedict Afful, and Renato A. Villano. 2017. 'Relationship between Financial Stress, Moonlighting and Teacher Attrition'. Working Paper nesra/wp/17/005. Accra: Network for Socioeconomic Research and Advancement. https://www.ssrn.com/abstract=2995561.

Krueger, Anne. 1974. 'The Political Economy of the Rent-Seeking Society'. *American Economic Review* 64 (3): 291–303.

Kwauk, Christina, and Jenny Perlman Robinson. 2016. 'Bridge International Academies Delivering Quality Education at a Low Cost in Kenya, Nigeria, and Uganda'. Washington, D.C.: Brookings Institute.

Mamdani, Mahmood. 1990. 'A Glimpse at African Studies, Made in USA'. *CODESRIA Bulletin*, no. 2: 7–11.

Mbiti, Isaac M. 2016. 'The Need for Accountability in Developing Countries'. *Journal of*

Economic Perspectives 30 (3): 109–32.
Mirowski, Philip, and Robert Van Horn. 2005. 'The Contract Research Organization and the Commercialization of Scientific Research'. *Social Studies of Science* 35 (4): 503–548.
Mkandawire, Thandika. 2000. 'Non-Organic Intellectuals and "Learning" in Policy-Making Africa'. *Learning in Development Co-Operation*, 205.
———. 2015. 'Neopatrimonialism and the Political Economy of Economic Performance in Africa: Critical Reflections'. *World Politics* 67 (3): 563–612.
Mkandawire, Thandika, and Charles C. Soludo. 1999. *Our Continent, Our Future: African Perspectives on Structural Adjustment*. Asmara: CODESRIA and Africa World Press.
Moffitt, Robert. 1992. 'Evaluation Methods for Program Entry Effects'. In *Evaluating Welfare and Training Programs*, edited by Charles Manski and Charles Garfinkel, 231–52. Cambridge, Massachusetts: Harvard University Press.
Mukudi, Edith. 2004. 'The Effects of User-Fee Policy on Attendance Rates among Kenyan Elementary Schoolchildren'. *International Review of Education / Internationale Zeitschrift Für Erziehungswissenschaft / Revue Internationale de l'Education* 50 (5/6): 447–61.
Mulkeen, Aidan. 2010. *Teachers in Anglophone Africa: Issues in Teacher Supply, Training, and Management*. Development Practice in Education. Washington, DC: World Bank.
Muller, Sean. 2014. *Randomised Trials for Policy: A Review of the External Validity of Treatment Effects*.
Muralidharan, Karthik, and Venkatesh Sundararaman. 2013. 'Contract Teachers: Experimental Evidence from India'. w19440. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w19440.
Mustapha, Abdul Raufu. 2012. 'The Public Sphere in 21st Century Africa: Broadening the Horizons of Democratisation'. *Africa Development / Afrique et Développement* 37 (1): 27–41.
Nishimura, Mikiko, and Takashi Yamano. 2013. 'Emerging Private Education in Africa: Determinants of School Choice in Rural Kenya'. *World Development* 43 (March): 266–75. https://doi.org/10.1016/j.worlddev.2012.10.001.
Norris, Susan L., Haley K. Holmer, Lauren A. Ogden, and Brittany U. Burda. 2011. 'Conflict of Interest in Clinical Practice Guideline Development: A Systematic Review'. *PloS One* 6 (10): e25153.
Oketch, Moses, and Caine Rolleston. 2007. 'Policies on Free Primary and Secondary Education in East Africa: Retrospect and Prospect'. *Review of Research in Education* 31: 131–58.
Olukoshi, Adebayo. 2006. 'African Scholars and African Studies'. *Development in Practice* 16 (6): 533–544.
Omwami, Edith Mukudi, and Edmond J. Keller. 2010. 'Public Funding and Budgetary Challenges to Providing Universal Access to Primary Education in Sub-Saharan Africa'. *International Review of Education / Internationale Zeitschrift Für Erziehungswissenschaft / Revue Internationale de l'Education* 56 (1): 5–31.
Osler, Audrey. 1997. 'Teachers' Biographies and Educational Development: A Kenyan Case Study'. *International Journal of Educational Development* 17 (4): 361–71. https://doi.org/10.1016/S0738-0593(97)00001-1.
Peters, Jörg, Jörg Langbein, and Gareth Roberts. 2016. 'Policy Evaluation, Randomized Controlled Trials, and External Validity—A Systematic Review'. *Economics Letters* 147 (October): 51–54. https://doi.org/10.1016/j.econlet.2016.08.013.
———. 2018. 'Generalization in the Tropics – Development Policy, Randomized Controlled Trials, and External Validity'. *The World Bank Research Observer* 33 (1): 34–64. https://doi.org/10.1093/wbro/lkx005.

Raudenbush, Stephen W. 2018. 'On Randomized Experimentation in Education: A Commentary on Deaton and Cartwright, in Honor of Frederick Mosteller'. *Social Science & Medicine* 210 (August): 63–66. https://doi.org/10.1016/j.socscimed.2018.04.030.

Ravallion, Martin. 2012. 'Fighting Poverty One Experiment at a Time: A Review of Abhijit Banerjee and Esther Duflo's *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*'. *Journal of Economic Literature* 50 (1): 103–14. https://doi.org/10.1257/jel.50.1.103.

Reddy, Sanjay. 2012. 'Randomise This! On Poor Economics'. *Review of Agrarian Studies* 2 (2): 60–73.

Romero, Mauricio, Justin Sandefur, and Wayne Aaron Sandholtz. 2018. 'Outsourcing Education in a Fragile State: Experimental Evidence from Liberia'. Working Paper.

Ron Balsera, Maria, Steven J. Klees, and David Archer. 2018. 'Financing Education: Why Should Tax Justice Be Part of the Solution?' *Compare: A Journal of Comparative and International Education* 48 (1): 147–62. https://doi.org/10.1080/03057925.2017.1394743.

Sayed, Yusuf, Azeem Badroodien, Yunus Omar, Eugene Ndabaga, Mario Novelli, Naureen Durrani, Angeline Barrett, Lorna Balie, Thomas Salmon, and Marcina Singh. 2018. 'Engaging Teachers in Peacebuilding in Post-Conflict Contexts: Rwanda and South Africa'. Synthesis report. Centre for International Education, Sussex University.

Stromquist, Nelly P. 1999. 'The Impact of Structural Adjustment Programmes in Africa and Latin America'. In *Gender, Education and Development: Beyond Access to Empowerment*, edited by Christine Heward and Sheila Bunwaree, 17–32. London: Zed Books.

Stromquist, Nelly P., Steven J. Klees, and Jing Lin. 2017. *Women Teachers in Africa: Challenges and Possibilities*. Routledge.

Tamale, Sylvia, and Joseph Oloka-Onyango. 1995. '" The Personal Is Political," or Why Women's Rights Are Indeed Human Rights: An African Perspective on International Feminism'. *Human Rights Quarterly* 17 (4): 691–731.

The Lancet. 2004. 'The World Bank Is Finally Embracing Science'. *The Lancet*, no. 364: 731–32.

Tilley, Helen. 2011. *Africa as a Living Laboratory: Empire, Development, and the Problem of Scientific Knowledge, 1870-1950*. Chicago: University of Chicago Press.

Tollison, Robert D. 1982. 'Rent Seeking: A Survey'. *Kyklos* 35 (4): 575–602. https://doi.org/10.1111/j.1467-6435.1982.tb00174.x.

Tullock, Gordon. 1967. 'The Welfare Costs of Tariffs, Monopolies, and Theft'. *Economic Inquiry* 5 (3): 224–232.

UK Government. 2012. 'Business: UK Impact Investments to Boost Development'. *UK Government* (blog). 7 December 2012. https://www. gov.uk/government/news/business-uk-impactinvestments- to-boost-development.

UKTI Education. 2011. 'About Us'. *UKTI Education* (blog). 2011. About us', https://www.gov.uk/government/organisations/ukti-education/about.

Vos, R., R. Bedi, P. Kimalu, D. Manda, and N. Nafula. 2004. 'Achieving Universal Primary Education: Can Kenya Afford It?' Working Paper. Department of Economics, University of Connecticut.

Wall Street Journal Editorial. 2017. 'The Real Democratic Party'. *Wall Street Journal*, 8 February 2017. https://www.wsj.com/articles/the-real-democratic-party-1486513439.

Walle, Nicolas van de. 2001. *African Economies and the Politics of Permanent Crisis, 1979-1999*. Cambridge: Cambridge University Press.

Wanzala, Ouma. 2018. 'TSC to Hire 88,000 New Teachers'. *Daily Nation*, 11 April 2018.

https://www.nation.co.ke/news/education/TSC-wins-MPs--approval-hire-88-000-teachers-/2643604-4381496-ukrm27z/index.html.
World Food Programme. 2012. 'Kenya Food Security and Outcome Monitoring'. Rome: World Food Programme. https://documents.wfp.org/stellent/groups/public/documents/ena/wfp255350.pdf.
World Vision. 2018. 'Our Partners'. 2018. https://www.worldvision.org.uk/who-we-are/our-partners/.
Zeleza, Paul Tiyambe. 2003. 'Academic Freedom in the Neo-Liberal Order: Governments, Globalization, Governance and Gender'. *Journal of Higher Education in Africa* 1 (1): 149–194.

Notes

[1] This figure should be treated with some caution and is likely an upper-bound estimate. Only 46% of the studies discussed the question of informed consent. And of these, 19% were unclear as to whether the participants gave consent to be part of an experiment, rather than gave consent to be part of a social programme.

[2] This data was gleaned from the appendix, which contained a list of all the papers reviewed. I identified the current institutional location of first authors by conducting a Google search, and then coded the institution by country and then region.

[3] The references in this article are from the authors' working paper from 2013, which was then published in 2018.

[4] Calculated using the World Bank's estimate of a PPP private consumption conversion factor of 35.43 for Kenya in 2011, accessed at: https://data.worldbank.org/indicator/PA.NUS.PRVT.PP?locations=KE.

[5] Put in international perspective, Kenya's GDP per capita is $2.71 per day in PPP dollars. A teacher would therefore earn 0.7 times the per capita GDP. In contrast, teachers earn 1.8 times per capita GDP in OECD countries, and 2.5 times per capita GDP in middle-income countries (Mulkeen 2010, 152).

[6] One might respond that lower wages for black people or women is a case of otherwise irrelevant identity markers being used as proxies of competence, such that they lead to inefficient hiring outcomes. On this view, labour market discrimination is defined by the misapplication of identity markers. Thus, differences in wages between black and white people count as a form of discrimination, but differences in wages between permanent and contract teachers do not. Instead, these differences are a form of efficiency improving measures, since they seek to bring down government wages to private sector levels. But consider a situation in which black workers in the public sector receive a higher wage than black workers in the private sector (as is the case in South Africa, see Bhorat et al. 2015). One might argue that black civil servants receive a higher wage because the government is compelled to do so by a largely black electorate and strong public sector unions, whereas black workers in the private sector continue to be subject to discrimination in the largely white-controlled private sector. In this scenario, setting private sector wages for black civil servants would therefore undermine democratic gains and reintroduce white racism into the state sector. Similarly, teachers in the Kenyan public system might receive higher wages because the state is compelled to pay salaries that meet minimum livelihood thresholds by democratic institutions, such as unions and courts, whereas the private sector is less encumbered by democratic oversight. The point is not that higher civil service wages are necessarily more just, but that further contextual evidence and theoretical argument is required to adjudicate between these different interpretations.

[7] Since the precise nature of this anti-empiricism is difficult to convey, the literature on neopatrimonialism often relies on "detailed descriptions and … imaginative portrayals of a local scene, replete with a kind of anthropological knowingness … [it] is full of vibrant metaphors and characterized by unbridled use of anecdotes, pejorative vocabulary … and vivid vignettes of the all-too-frequent cases of egregious abuse of state resources and power. (Mkandawire 2015, 564) While this may strengthen the rhetorical force of the argument, it does nothing to justify its empirical validity.

[8] The concept of a "living laboratory" has been used to understand the kinds of scientific knowledge used in British colonial administration (Tilley 2011). The term is drawn from Lord Hailey's opening remarks from

the African Research Survey (1929-1938) which was conducted in an effort to better understand, govern and 'develop' African societies in the tropics. "Africa presents itself as a living laboratory", Hailey wrote, "in which the reward of study may prove to be not merely the satisfaction of an intellectual impulse, but an effective addition to the welfare of the people." (Hailey 1938, xxiv–xv) However, the allied concept of *champs d'experience*, or experimental terrains, has also been applied in the context of French colonial rule, especially in the Maghreb (Celik 1997).